A Quantification Model of Grammaticality

Philippe Blache, Jean-Philippe Prost

A Quantification Model of Grammaticality

Philippe Blache[1] and Jean-Philippe Prost[1 2]

[1] Laboratoire Parole & Langage, CNRS, Université de Provence,
29 Avenue Robert Schuman, 13621 Aix-en-Provence, France
blache@lpl-aix.fr
[2] Centre for Language Technology, Macquarie University
Sydney NSW 2109, Australia
jean-philippe.prost@lpl-aix.fr

Abstract. The traditional binary notion of *grammaticality* is more and more often replaced by intermediate levels of acceptability, also called *gradience*. This paper aims to provide a numerical account of syntactic gradience. It introduces and investigates a numerical model with which acceptability can be predicted by factors derivable from the output of a parser. Its performance is compared to other experiments, and the fit of each model is evaluated. Our model shows a good correlation with human judgement of acceptability.

1 Introduction

Grammaticality is not binary (see [Chomsky75]) but rather a scale phenomenon: some sentences are judged as more grammatical than some others. This question has been recently described in terms of *gradience* ([Pullum01], [Fanselow06], [Aarts07]) and replaces *grammaticality* by intermediate levels of acceptability. Gradience phenomena have different sources. One is the violation of rules or constraints of a given language. These rules playing a more or less important role in the linguistic structure, it is possible to grade them, as it is done in probabilistic approaches (see [Manning99]) or, in a more theoretical perspective, in the Optimality Theory ([Prince93]). As a side effect, grading the syntactic properties makes it possible to explain the differences in acceptability judgements by human. For example, the repetition of a determiner as in (1a) can be considered as less problematic than a problem of word order as in (1b):

(1) a. *Buildings burn in the the Kenyan town of Eldoret.*
    b. *The adopted board the regulations.*

Another source of gradience in the evaluation of grammaticality comes from the syntactic structure itself. A same constraint violation is more or less acceptable according to its context or its location in the structure. For example, (2b) seems to be more acceptable than (2a). In both cases, the same linearity constraint is violated. The difference comes from the fact that the constituent affected by the violation is more deeply embedded in (2b) than in (2a). In the second sentence, the acceptability judgment is more affected by parsing difficulty than constraint violation.

(2) a. *Paul phoned father his.*
b. *Paul came back into the house that was built by father his.*

This example illustrates the fact that a violation can be, to a certain extent, balanced by the context, or more generally by information born by other domains, including prosody or pragmatics.

In this paper we propose a model for syntactic gradience,which enables quantifying the phenomenon. This paper is organised in three parts: we first propose to make precise the notion of gradience, its theoretical description and what are the needs for its quantification. In the second section we introduce a model, which is a refinement of the one described in [Blache06a]. In the last part, we validate the model using a parser's output. The results are evaluated first by comparison with those of a psycholinguistic experiment, then over a large corpus of unrestricted text.

2 Linear Optimality Theory

Several experiments have shown the effect of constraint violation on grammaticality. In particular, preliminary works by [Legendre90] (anticipating Optimality Theory) have proposed to quantify this effect. More recently, [Keller00] explored this question through in-depth description of different constructions. This work showed in particular the cumulativity effect, that led to the *Linearity Hypothesis* indicating that the "*grammaticality of a structure is proportional to the weighted sum of the contraint violations it incurs*". Starting from this hypothesis, Keller elaborates the *Linear Optimality Theory* (see [Keller06]). Linearity makes it possible to give to the notion of harmony a particular definition: the harmony of a structure it the opposite of the sum of the violated constraint weights. The grammaticality of different structures can be then evaluated and compared. The following figure, taken from [Keller06], illustrate the process. It shows 4 different candidate structures (S_1 to S_4) and three different constraints (C_1, C_2, C_3), with the respective weights 4, 3, and 1. Constraint violations are indicated by *. The figure gives the harmony of the different structures.

	C_1	C_2	C_3	
Structure	4	3	1	*Harmony*
S_1		*	*	-4
S_2		*	**	-5
S_3			*	-1
S_4	*			-4

Fig. 1. Structures, constraint violations and harmony

This formal examples illustrates the cumulativity effect: the multiple constraint violation of the structure S_2 makes its harmony worst than that of S_4

in spite of the fact that this last structure violates a strongest constraint. Moreover, the harmony values show that S_1 and S_4 are at the same level in terms of violation weights, in spite of the fact that they violate different types and number of constraints.

The structure hierarchy is then $S_3 > \{S_1, S_4\} > S_2$ (where the hierarchy in classical OT would be $S_3 > S_1 > S_2 > S_4$). Several experiments (see [Keller00]) has shown the validity of the linearity hypothesis.

However, several questions can not be addressed in this framework. First, this view of cumulativity is purely quantitative and does not take into consideration the structure itself. Concretely, it would not predict any acceptability variation between (2a) and (2b).

Second, OT was designed in order to compare several structures by means of constraint violation. Subsequently, two structures violating the same constraint can not be discriminated. The following example illustrates this phenomenon. Both *NP* violate the same constraint (gender agreement violation between the noun and the past participle). However, the same violation embedded into a relative clause in (3b) renders the NP more acceptable than (3b).

(3) a. *La maison détruit par l'explosion*
The house-fem destroy-PPast-masc by the explosion
b. *La maison qui a été détruit pas l'explosion*
The house-fem that has been destroy-PPast-masc by the explosion

Last, and more importantly, it is important to compare utterances that do not violate any constraint without being at the same level in terms of acceptability. In the following examples, the different sentences do not bring the same quantity of information. The different constructions instantiate differently the direct object. The first sentence is a cleft, there is a strict an unambiguous interpretation of the *NP* cleft as a direct object, the function being marked by the accusative mark of the relative pronoun "*que*". The example (5b) is a dislocation with a resumptive accusative pronoun "*la*", possibly referring to the dislocated *NP*. Finally, the last sentence is also a kind of dislocation, without any coreference phenomenon, the status of the extracted *NP* being ambiguous (direct object, vocative, etc.)

(4) a. *C'est Marie que je supporte pas.*
It is Mary that-Acc I can't stand.
b. *Marie je la supporte pas*
Mary I can't stand her.
c. *Marie je supporte pas.*
Mary I can't stand.

This example are graded: the first example is unambiguous because of the weight of its syntactic information: several morpho-syntactic and syntactic constraints are satisfied there. The second example contains less information, it is ambiguous (dislocation vs. vocative interpretation). In this case, the only visible relation concerns the possible agreement between the NP and the embedded clitic. Finally, the last example is even more ambiguous: Marie can receive a vocative, an accusative or even a dative interpretation there, without any morphology or syntactic mark.

These last examples show the necessity of taking into account, on top of the violated constraints, also the satisfied one: the quantity and the weight of satisfied constraints can play a role, as the violated on, in the ranking process.

3 Property Grammars

The Property Grammars (PG) approach [Blache05a] is purely constraint-based: all syntactic information is represented by means of constraints; no external device such as the *Gen* function in OT or the generation of the dependency network in Constraint Dependency Grammar [Maruyama90] is required. Several constraint types are used: *Constituency, Uniqueness, Precedence, Obligation, Requirement* and *Exclusion*. They can be specified in terms of constraints over directed graphs, as presented in figure (2).

- $Const(A, B) : (\forall x, y)[(A(x) \wedge B(y) \rightarrow x \triangleleft y]$
Classical definition of constituency, represented by the dominance relation indicating that a category B is constituent of A.

- $Uniq(A) : (\forall x, y)[A(x) \wedge A(y) \rightarrow x \approx y]$
If one node of category A is realized, there cannot exists other nodes with the same category A. Uniqueness stipulates constituents that cannot be repeated in a given construction.

- $Prec(A, B) : (\forall x, y)[(A(x) \wedge B(y) \rightarrow y \not\prec x)]$
This is the linear precedence relation: if the nodes x and y are realized, then y cannot precedes x

- $Oblig(A) : (\exists x)(\forall y)[A(x) \wedge A(y) \rightarrow x \approx y]$
There exists a node x of category A and there is no other node y of the same category. An obligatory category is realized exactly once.

- $Req(A, B) : (\forall x, y)[A(x) \rightarrow B(y)]$
If a node x of category A is realized, a node y of category B has too. This relation implements cooccurrence restrictions.

- $Excl(A, B) : (\forall x)(\nexists y)[A(x) \wedge B(y)]$
When x exists, there cannot exist a sibling y. This is the exclusion relation between two constituents.

Fig. 2. *Constraint types in PG*

The following example illustrates some constraints describing the *NP* and the *AP* (in which a dependency constraint has been added):

`Cx_NP` : *Prec(Det, N)* $\wedge$ *Oblig(N)* $\wedge$ *Req(N, Det)* $\wedge$ *Excl(N, Pro)* $\wedge$ *Dep(Det, N)* $\wedge$ *Dep(AP, N)*

`Cx_AP` : *Prec(Adv, Adj)* $\wedge$ *Oblig(Adj)* $\wedge$ *Dep(Adv, Adj)*

A grammar is then a set of constraints. Parsing an input consists of evaluating this constraint system for a given assignment (i.e. the set of categories corresponding to the input words). The outcome is a description of the input, which is made of the set of the evaluated constraints. Depending on the form of the input this set may contain both satisfied and violated constraints.

In PG constraints thus play a double role (as it is the case in constraint programming): they can rule out structures as well as instantiate values. Such an approach is well-suited to modeling gradience because the constraints are independent from one another. Indeed the basic assumptions in PG, unlike in OT, stipulate that constraints are *unranked*, *local* and *violable*:

Unranked: A limited set of constraint types is used, each one with its own operational semantics. A constraint bears information of a unique type, representing a single and homogeneous piece. A constraint is thus atomic and can be evaluated independently from the other ones. Subsequently, by being mutually independent and never specified with respect to others. The weighting mechanism used in PG is not an order relation over the set of constraints, but a property specific to each constraint.

Violable: Allowing constraint violation is a pre-requisite when dealing with partial or non-canonical inputs. In PG, all constraints are violable, but not necessarily. In OT, as in *Weighted CDG* (see [Schröder02]), the number and the type of constraint violations is used to compare two structures. Violation is then necessary, and no optimal candidate can be selected when all constraints are satisfied. In PG, constraint violation is not required; it only introduces flexibility.

Local: Universality is not a required property of PG. On the contrary, all constraints are local to a construction. It is an important difference with OT, not only theoretically, but also in the way of designing and using constraints. As indicated above, OT can only compare two structures with respect to constraint violation. It explains the fact that constraints must be specified in a very general and imperative way: the more general a constraint is, the more frequently it is violated. Universality must then be understood in OT as a mechanism which favors constraint violation. In PG, no universality is required; constraints can be specified at any level.

4 A Computational Model for Syntactic Gradience

The existing accounts of gradience previously mentioned rely on several properties:

- *Constraint violation*: this is a pre-requisite. Constraints must be be defeasible in order to describe any kind of input, whatever its form.
- *Constraint weighting*: a comparison necessarily relies on the possibility of measuring the impact of the different constraints (see [Foth05]).
- *Cumulativity*: the effect of constraint violation was shown to be cumulative [Keller00].

However, these properties do not capture all phenomena that have been presented in the previous section. First, we have to explain the effects of a relative counterbalance between satisfied and violated constraints: constraint violation can be in some cases attenuated by the importance of satisfied constraints. Moreover the form of the syntactic structure (flat or deep) and the location of constraint violation in the structure also have importance (cf. examples 3 and 4). Thus we propose to complete the list of properties needed to describe gradience by the following ones:

- *Constraint counterbalance*: cumulativity must take into account both violated and satisfied constraints.
- *Violation position*: the embedded level of the violation site in the syntactic structure carries consequences on acceptability.

An approach integrating these different properties offers, on top of a precise description of gradience, the possibility of calculating an index for any input, and quantifying, at least partially, its grammaticality. In the next section we detail and organise the basic information on top of which our model is built.

4.1 Basic information

in this section we present, in the form of different postulates, the kind of information required to build a computational model of gradience.

Failure Cumulativity As in other approaches, we postulate that acceptability is impacted by the amount of constraints it violates. We note N_c^- the amount of constraints violated by the constituent c (this factor corresponds to cumulativity in LOT).

Success Cumulativity Gradience is also affected by successful constraints. That is, an utterance acceptability is impacted by the amount of constraints it satisfies. We postulate that some form of interaction between satisfied and violated constraints contributes to a gradient of acceptability. We note N_c^+ the amount of constraints satisfied by the constituent c, and $E_c = N_c^+ + N_c^-$.

Constraint Weighting We postulate that constraints are weighted according to their influence on acceptability. The question of whether such weights are proportional to the importance of either constraint success or failure is addressed in assuming that a given constraint is of same relative importance either way in absolute value. We note W_c^+ (respectively W_c^-) the sum of the weights assigned to the constraints satisfied (respectively violated) by the constituent c.

A weight may be of different *scope* and *granularity*. The *scope* has to do with how widely a weight applies (to a constraint type or to an individual constraint). *Granularity* concerns the level a weight applies at (the grammar vs. the construction level). Scope and granularity can then be combined in different ways: all constraints from the same type at the grammar level, or all constraints from the same type at the constituent level, or individual constraints at the constituent level, or individual constraints at the grammar level—the difference between the last two possibilities assuming that a same constraint may occur in the specification of more than one category. Although the more fine-grained and the narrower the scope, the more flexible and accurate the influence on gradience, a too fine granularity and a too narrow scope (as in [Schröder02]) are also quite complex to manage. Therefore, we opted for a fair compromise, where the weighting scheme is restricted to the constraint types at the grammar level, which means that all constraints from the same type in the grammar are assigned same weight. For examples, all constraints of linearity (i.e. word order) are weighted 20, all constraints of obligation (i.e. heads) are weighted 10, and so on.

Constructional Density Acceptability is also impacted by the *density* of the constituent structure (*i.e.*, the quantity of information born by the structure). This notion is measured by the amount of constraints specifying a category in the grammar. We note $T_{\mathcal{C}}$ the total amount of constraints specifying the category $\mathcal{C}$ of the constituent c. The underlying idea is to balance constraint violations by the amount of specified constraints: without such a precaution one violation in a rather non-complex construction, such as AP—only specified by 7 constraints in our grammar—would be proportionally much more costly than one violation in a rather complex construction, such as NP—specified by 14 constraints.

Propagation Acceptability also depends on that of its nested constituents. Therefore, we postulate that acceptability is propagated in the constituent structure through the relationship of dominance. We note Z_c the number of nested constituents in c.

Constraint violation affects both realization, but in a lesser extent when the violation is deeply embedded (example 5b). Subsequently the models we investigate are recursive functions of their constituents' score.

4.2 Rating Models

A rating model for gradience aims to place an item along a scale by assigning it a score (rate).

Scoring Terms The various scoring components presented here aim to capture the factors postulated above. Each component is meaningful as such, but not sufficient when considered alone.

Satisfaction/Violation Ratio The *SRatio* ϱ_c^+ (resp. VRatio ϱ_c^-) is defined for the constituent c as follows:

$$\varrho_c^+ = \frac{N_c^+}{E_c} \qquad \varrho_c^- = \frac{N_c^-}{E_c}$$

The SRatio and violation ratio (VRatio) capture the postulates of Success and Failure Cumulativity respectively.

Completeness Index

The *Index of Completeness* for the constituent c of category $\mathcal{C}$ is defined as the following ratio, T being the total number of constraints describing the category in the grammar:

$$\mathcal{E}_c = \frac{E_c}{T_{\mathcal{C}}}$$

This score contributes to implement the postulate of Constructional Density, which suggests that the complexity of a constituent influences its acceptability.

Quality Index

The *Index of Quality* for the constituent c is defined as the following ratio:

$$\mathcal{W}_c = \frac{W^+ - W^-}{W^+ + W^-}$$

The *quality* of a constituent implements the postulate of Constraint Weighting, which suggests that all constraints do not have same importance with respect to acceptability, and therefore must be weighted accordingly.

Precision Index

The *Index of Precision* for the constituent c is defined as the following ratio:

$$\mathcal{P}_c = k \cdot \mathcal{W}_c + l \cdot \varrho_c^+ + m \cdot \mathcal{E}_c$$

These *adjustment coefficients* (k, l, m) are used as variable parameters for tuning up the model.

We observed that the SRatio in use in the Precision score seems to overemphasise the role of success cumulativity, that is, the role of the successful constraints characterising a constituent. Therefore, we define an index of *anti-precision*, where the SRatio term in the precision index is replaced by the VRatio as a negative term. *Anti-Precision Index*

We define the *Index of Anti-Precision* for the constituent c as the following ratio:

$$\widetilde{\mathcal{P}}_c = k \cdot \mathcal{W}_c - l \cdot \varrho_c^- + m \cdot \mathcal{E}_c$$

Compared to the precision score, the anti-precision rather emphasises the factor of Failure Cumulativity.

Rating Functions A rating function combines different scoring terms into a single score. Among the numerous functions investigated, the following ones more particularly draw our attention for the significance of their results. Grammaticality Index

The *Index of Grammaticality* (g) for the constituent c is defined recursively as follows (where c_i is a nested constituent of c):

$$g_c = \mathcal{P}_c \cdot \overline{g_{c_i}} = \mathcal{P}_c \cdot \frac{\sum_{i=1}^{Z_c} g_{c_i}}{Z_c}$$

Next to the g–model we define below a new model, in order to run a comparative investigation of the two. The index of *coherence* is similar to the one of grammaticality, except that it relies on anti-precision rather than precision.

Coherence We define the *Coherence* of a constituent c recursively as follows:

$$\gamma_c = \widetilde{\mathcal{P}}_c \cdot \overline{\gamma_{c_i}} = \widetilde{\mathcal{P}}_c \cdot \frac{\sum_{i=1}^{Z_c} \gamma_{c_i}}{Z_c}$$

A comparison with LOT We have seen that cumulativity makes it possible in LOT to rank the structures with respect to their constraint violation. In the figure (3), we can compare the ranking by LOT with the obtained by our model. We can see that our ranking makes it possible to precise the result of LOT in providing an intermediate ranking between S_1 and S_4. As expected, S_4 satisfying more constraints than S_1, its precision index is higher.

	C_1	C_2	C_3		
Structure	4	3	1	*Harmony*	Precision index
S_1		*	*	-4	0.537
S_2		*	**	-5	0.379
S_3			*	-1	0.842
S_4	*			-4	0.620

Fig. 3. Structures, constraint violations and harmony

The following table recapitulates the rankings obtained by the different models (OT, LOT and GP) :

$$
\begin{array}{ll}
\text{OT:} & S_3 > S_1 > S_2 > S_4 \\
\text{LOT:} & S_3 > \{S_1, S_4\} > S_2 \\
\text{GP:} & S_3 > S_4 > S_1 > S_2
\end{array}
$$

5 Experimental Validation

We investigate to what extent the models of syntactic gradience presented above fit acceptability judgement by human standards.

5.1 Psycholinguistics experiment

[Blache06a] reports an experiment set up with psycholinguists. It shows a correlation between the Grammaticality Index (GI) (γ–model, above) and acceptability judgements provided by subjects.

The experiment relies on a set of sentences in which constraint violation was controlled. 20 types of sentence were designed, in which at most two constraints are violated. Several base sentences were created, each one generating the 20 types. As a result, 60 sentences from the different types were presented for evaluation to 44 subjects. The subjects were asked to rate the sentences, using *Magnitude Estimation* (see [Bard96]).

Next to this evaluation, GIs were calculated semi-automatically for each sentence: a generic syntactic structure (*i.e.* a syntactic tree) was associated to the phrase types, together with its characterisation (*i.e.* the set of satisfied and violated constraints) of each constituent. Figure (2) shows example sentences along with their GI.

The subjects' judgements were then compared to a rating relying on grammaticality indexes. A very good correlation (coefficient $\rho_1 = 0.76$) was observed between GI and acceptability judgement. An even better correlation with a coefficient $\rho_2 = 0.87$ is reported on a smaller sample of corrected data.

5.2 Computational Validation

For the work we present here we have experimented the two models (g and γ) in using the output from parsers based on PG. We have tested two different parsers, both robust: the first one is a chart parser using dynamic programming [Prost06], and is a direct interpretation of PG. It explores the entire search space, and relies on constraint satisfiability in order to build the set of structures. The second parser [Blache06b] is non-deterministic and relies on control heuristics consisting in selecting construction types by means of precedence and constituency constraints (corresponding to a left-corner like strategy). The parsers using different techniques, strategies and level of analyses, they may build different solutions for the same input (especially due to the non-determinism of one of them).

Both parsers show comparable results with respect to gradience quantification (described in this section). The robust parser has been used in the evaluation over the large corpus (next section). The first evaluation, presented here, consists of replicating automatically the experiment described in the previous section. The goal is to show a correlation between the predictions from the model and the subjects' assessment.

We progressively tune up the models by assigning values to the different parameters (i.e. adjustment coefficients and constraint weights). The problem consists of finding out the right order of magnitude among the different parameters in order to obtain the best possible correlation with the values of acceptability. Different combinations were attempted. A sample of the correlations obtained is reported in table 1. The best correlation ($\rho = 0.5425$) is obtained for record

No violations	
11. Marie *a emprunté un très long chemin pour le retour*	0.465
NP-violations	
21. *Marie a emprunté très long chemin un pour le retour*	-0.643
22. *Marie a emprunté un très long chemin chemin pour le retour*	-0.161
...	
VP-violations	
51. *Marie un très long chemin a emprunté pour le retour*	-0.56 *
54. *Marie emprunté un très long chemin pour le retour*	-0.322 *
...	

Fig. 4. Acceptability Results

$\sharp$	Adjust.			Weight						Correlation		
	k	l	m	wl	wo	we	wr	wd	wu	Max	g	γ
8	4	2	1	20	3	5	4	2	10	0.4658	0.3932	$^\natural$**0.4658**
11	4	2	1	5	3	2	2	0	2	0.4945	0.3891	$^\natural$**0.4945**
12	4	2	1	5	3	2	2	1	2	0.4946	0.3805	$^\natural$**0.4946**
17	4	2	1	20	10	5	4	3	2	0.5425	0.4745	$^\natural$**0.5425**
										0.5425		

Table 1. Calibration of adjustments and constraint weights. *The weights are those assigned to the different constraint types: Linearity (wl), Obligation (wo), Exclusion (we), Requirement (wr), Dependency (wd), and Uniqueness (wu); col.* $\sharp$ *is a record Id; col.* Max *contains the maximum correlation for each record.*

$\sharp$17, for the γ–model. The scatter plot from fig. 5 illustrates how the γ–model fits acceptability judgement.

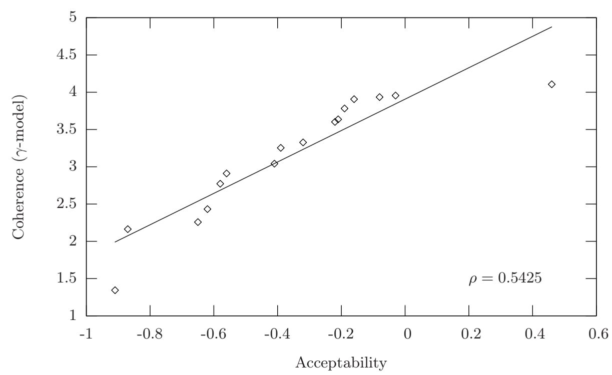


Fig. 5. Correlation Acceptability vs. Coherence

The constraint weights are ranked $wu < wd < wr < we < wo < wl$. It is important to emphasise that unlike in [Keller00], the constraints are not only ranked according to how much unacceptability they entail, but according to how important they are in *absolute value* with respect to acceptability.

Whatever the combination of parameters, γ always outperforms g, which confirms that this latter model is over-emphasising the role success cumulativity compared to the role of failure cumulativity.

The best performing scheme of parameters (rec. $\sharp$17) grants a great deal of importance to Linearity (a factor 10 between $wl = 20$ and the minimum $wu = 2$, and a factor 2 between wl and its next follower $wo = 10$), then to Obligation (a factor 5 between $wo = 10$ and wu, and a factor 2 between wo and it very next follower $we = 5$). Then follow the remaining weights, ranging over $[2 \ldots 5]$. This observation of two constraint types (namely Linearity and Obligation) on one hand, and the other ones on the other hand, tends to confirm the hard vs. soft dichotomy discussed by Keller.

Reduced data sample from the psycholinguistics experiment:

In order to perform a more accurate comparison between our results and that reported in [Blache06a], we ran a series of experiments using the same data sample, which is a subset of the full corpus. The results are reported in table 2.

♯	Adjust.			Weight						Correlation		
	k	l	m	wl	wo	we	wr	wd	wu	Max	g	γ
2	4	2	1	5	3	2	2	0	2	0.6017	0.5408	$^\natural$**0.6017**
3	4	2	1	5	3	2	2	1	2	0.6017	0.5246	$^\natural$**0.6017**
4	4	2	1	20	10	5	4	3	2	0.6427	$^\natural$**0.6427**	0.6024
										0.6427		

Table 2. Correlations on the reduced data sample.

The best correlation (rec. ♯4) is obtained for the same parameter scheme as the best one from table 1, but surprisingly this time g outperforms the other two models. It confirms the crucial influence of Linearity on acceptability, but the roles of Uniqueness and Obligation is still unclear, though they are seemingly preponderant.

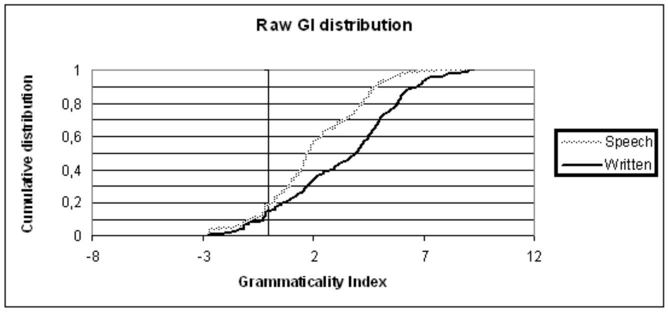
This first evaluation confirms the validity of our gradience model in showing a strong correlation with subjects predictions. Moreover, and this is the most important result, these predictions have been obtained automatically (at the difference with [Blache06a]), which opens the possibility of large scale experimentation, as presented in the next section.

5.3 Large Corpus Validation

We ran our model on a large French corpus, made from different sources: newspapers (184,367 words), spontaneous spoken language (14,065 words) and radio broadcasts transcriptions (84,685 words).

Some general figures can be given. Figure (1) illustrates the (unsurprising) difference between oral and written text: we observe that a greater proportion of sentences in oral production with low grammaticality index. As illustrated in figure (1), the mean index for spoken corpora is 2.46 whereas that of written is 3.36. What is more interesting is the repartition of the index values. There are only few differences between the two types: low and high index values are very similar, which means that some written sentences have a very low grammaticality index whereas some spoken (including spontaneous) can be very high.

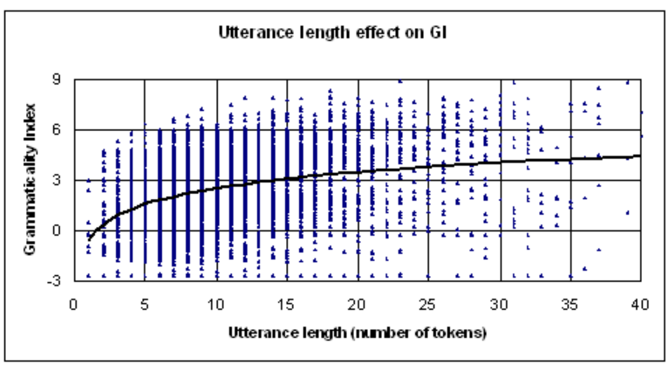
As for our postulates, these results confirm the relevance of success cumulativity. This effect is illustrated in the following examples (from the corpus): the longer sentence (6) obtains a higher score than the short one.



(5) j'aime la cuisine (2,1385)
I like cooking
(6) donc c'est pas évident parce qu'il y a des jours où il y a pas de boulot il y a des jours où il y a du boulot comme partout (6,7266)

So it is not easy because there are days where there is no work there are days where there is work, like everywhere

A the corpus scale, we can illustrate this phenomenon in correlating grammaticality index with sentence size. The following figure shows the repartition of sentences with respect to the size. The average of the indexes in function of size match the curve $\langle\rho\rangle = \ln(size)$.



In other words, as expected, the number of words can increase the grammaticality index level. Again, as observed in other works, cumulativity remains less important than the importance (i.e. the weight) of violated constraints. This aspect is illustrated in the following examples: a precedence constraint has been violated in (8), explaining the lower score than in (7):

(7) des foyers genre foyers ce qu'on appelle foyers de jeunes filles ou non mixte quoi (0,8089)

Boarding houses sort of boarding houses what is called boarding houses for girls or not mixed like

(8) non ça m'a fait vraiment pas mal cogiter mais mais bien quoi c'est (-1,201)

No it really made me think hard but but right that's

Other assumptions can be confirmed by a detailed analysis of the results. At this level, we can see that most of constraint violation concern uniqueness, requirement and obligation.

6 Conclusion

We have propose in this paper a specification of the needs for a precise account of syntactic gradience. On this basis, we have specified a new computational model, taking advantage of a fully constraint-based syntactic representation as proposed in Property Grammars. This model has been evaluated first in replicating automatically a previous experiment showing the correlation between the scores given by the model and subjects acceptability judgements. We have then validate the approach in experimenting it on larger and unrestricted corpora.

An automatic account of gradience, such as the one presented here, can have many applications. In terms of parsing, it constitutes an efficient heuristic, helping in the selection of the construction types. Other applications can be imagined, for example in second language learning systems, helping the user in evaluating its productions. At a theoretical and cognitive level, this model shows the relevance of constraints in modelling language production and perception: the rating functions can help in explaining sentence complexity.

References

[Aarts04] Aarts B. (2004) "Modelling Linguistic Gradience", *Studies in Language*, 28(1):1–49.

[Aarts07] Aarts. B (2007) *Syntactic Gradience. The nature of Grammatical Indeterminacy*, Oxford University Press.

[Bard96] Bard E., D. Robertson & A. Sorace (1996) "Magnitude Estimation of Linguistic Acceptability", Language 72:1.

[Blache05a] Blache P. & J.-P. Prost (2005) "Gradience, Constructions and Constraint Systems", in H. Christiansen & al. (eds), *Constraint Solving and NLP*, LNCS, Springer.

[Blache05b] Blache P. (2005) "Property Grammars: A Fully Constraint-Based Theory", in H. Christiansen & al. (eds), *Constraint Solving and NLP*, LNCS, Springer.

[Blache06a] Blache P., B. Hemforth & S. Rauzy (2006) "Acceptability Prediction by Means of Grammaticality Quantification" , in proceedings of *COLING-ACL 06*

[Blache06b] Blache P. (2006) "A Robust and Efficient Parser for Non-Canonical Inputs", in proceedings of *ROMAND-06*,

[Bresnan01] Bresnan J., S. Dingare & C. Manning (2001) "Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi", in proceedings of *LFG'01*.

[Chomsky75] Chomsky N. (1975) *The Logical Structure of Linguistic Theory*, Plenum press.

[Fanselow06] Fanselow G., C. Féry, R. Vogel & M. Schlesewsky (eds.), *Gradience in Grammar: Generative Perspectives*, Oxford University Press.

[Foth05] Foth K., M. Daum & W. Menzel (2005) "Parsing Unrestricted German Text with Defeasible Constraints", in H. Christiansen & al. (eds), *Constraint Solving and NLP*, LNCS, Springer.

[Fouvry2003] Fouvry F. (2003), *Robust processing for constraint-based grammar formalisms*, Ph.D. thesis, Graduate School, University of Essex, Colchester, United Kingdom, April.

[Haji-Abdolhosseini05] Haji-Abdolhosseini M. (2005) *Modularity And Soft Constraints: A Study of Conflict Resolution in Grammar*, Ph.D. thesis, University of Toronto.

[Keller00] Keller F. (2000) *Gradience in Grammar. Experimental and Computational Aspects of Degrees of Grammaticality*, Phd Thesis, University of Edinburgh.

[Keller06] Keller F. (2006) "Linear Optimality Theory as a Model of Gradience in Grammar", in Gisbert Fanselow, Caroline Féry, Ralph Vogel, and Matthias Schlesewsky (eds.), *Gradience in Grammar: Generative Perspectives*, Oxford University Press.

[Legendre90] Legendre G., Y. Miyata & P. Smolensky (1990) "Harmonic Grammar - A Formal Multi-level Connectionist Theory of Linguistic Well-formedness: Theoretical Foundations", in *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*.

[Legendre06] Legendre G., A. Sorace & P. Smolensky (2006), "The Optimality Theory-Harmonic Grammar connection", in P. Smolensky & G. Legendre (eds.) *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar*, MIT Press.

[Manning99] Manning C. & H. Schütze (1999), Foundations of Statistical Natural Language Processing, MIT Press.

[Maruyama90] Maruyama H. (1990), "Structural Disambiguation with Constraint Propagation", in proceedings of *ACL'90*.

[Menzel98] Menzel W. & I. Schroder (1998) "Decision procedures for dependency parsing using graded constraints", in proceedings of *ColingACL Workshop on Processing of Dependency-based Grammars*.

[Prost06] Prost J-Ph. (2006) "Numbat: Abolishing Privileges when Licensing New Constituents in Constraint-Oriented Parsing", in proceedings of *CSLP-06*.

[Prince93] Prince A. & Smolensky P. (1993) *Optimality Theory: Constraint Interaction in Generative Grammars*, Technical Report RUCCS TR-2, Rutgers Center for Cognitive Science.

[Pullum01] Pullum G. & B. Scholz (2001) "On the distinction between model-theoretic and generative-enumerative syntactic frameworks", in processdings of *LACL-01*.

[Schröder02] Schröder I. (2002) *Natural Language Parsing with Graded Constraints*. PhD Thesis, University of Hamburg.